

# Population Genomics and Inference of *Mycobacterium avium* Complex Clusters in Cystic Fibrosis Care Centers, United States

Nabeeh A. Hasan,<sup>1</sup> Rebecca M. Davidson,<sup>1</sup> L. Elaine Epperson, Sara M. Kammlade, Sean Beagle, Adrah R. Levin, Vinicius Calado de Moura, Joshua J. Hunkins, Natalia Weakly, Scott D. Sagel, Stacey L. Martiniano, Max Salfinger, Charles L. Daley, Jerry A. Nick, Michael Strong

*Mycobacterium avium* complex (MAC) species constitute most mycobacteria infections in persons with cystic fibrosis (CF) in the United States, but little is known about their genomic diversity or transmission. During 2016–2020, we performed whole-genome sequencing on 364 MAC isolates from 186 persons with CF from 42 cystic fibrosis care centers (CFCCs) across 23 states. We compared isolate genomes to identify instances of shared strains between persons with CF. Among persons with multiple isolates sequenced, 15/56 (27%) had >1 MAC strain type. Genomic comparisons revealed 18 clusters of highly similar isolates; 8 of these clusters had patients who shared CFCCs, which included 27/186 (15%) persons with CF. We provide genomic evidence of highly similar MAC strains shared among patients at the same CFCCs. Polyclonal infections and high genetic similarity between MAC isolates are consistent with multiple modes of acquisition for persons with CF to acquire MAC infections.

**N**ontuberculous mycobacteria (NTM) are ubiquitous microorganisms found in indoor and outdoor habitats, including water, soil, and dust. NTM can infect susceptible persons, including those with lung diseases such as cystic fibrosis (CF) (1). Previous surveys conducted in the United States have found that *Mycobacterium avium* complex (MAC) species are

clinically relevant and the most frequently isolated NTM (2). MAC consists of 9 slow-growing mycobacterial species (3–6), of which the 2 most frequently observed are *M. avium* (MAV) and *M. intracellulare*, including its subspecies *intracellulare* (MINT) and subspecies *chimaera* (MCHIM) (4). In the United States, most persons with CF and positive NTM cultures (61%) had MAC species infections (2,7). MAC infections increased by 3% annually during 2010–2014.

MAC pulmonary infections are probably acquired by inhalation of aerosols (8), but the sources and modes of transmission of MAC remain unclear. Studies using various molecular genotyping methods have shown MAC isolates from human airway samples to have high genetic similarity to isolates from animals (8–10), water (11,12), bathroom faucets (13), showerheads (14,15), pools (16), and soil (17). Other potential MAC infection sources include fomites, zoonotic sources, and contaminated materials (10,18). Despite the clinical relevance of MAC and its prevalence among persons with CF, the genomic relationships of MAC isolates and the potential for person-to-person transmission are poorly understood. Whole-genome sequencing (WGS) to analyze the genetic diversity of MAC is aimed at identifying MAC infections that cluster by high bacterial genomic sequence similarity, particularly in susceptible populations such as persons with CF. Unclustered isolates are unrelated and are therefore not implicated in transmission, but clustering between MAC isolates suggests that they are derived from the same source (i.e., shared water, surfaces, or person-to-person transmission). To this end, we analyzed the WGS of NTM isolates

Author affiliations: National Jewish Health, Denver, Colorado, USA (N.A. Hasan, R.M. Davidson, L.E. Epperson, S.M. Kammlade, S. Beagle, A.R. Levin, V. Calado de Moura, J.J. Hunkins, N. Weakly, C.L. Daley, J.A. Nick, M. Strong); University of Colorado Anschutz Medical Campus and Children's Hospital Colorado, Aurora, Colorado, USA (S.D. Sagel, S.L. Martiniano); University of South Florida College of Public Health and Morsani College of Medicine, Tampa, Florida, USA (M. Salfinger)

DOI: <https://doi.org/10.3201/eid2711.210124>

<sup>1</sup>These first authors contributed equally to this article.

voluntarily sent from US CF care centers (CFCCs) during a 4-year period. The goals of this project were to support routine clinical care through high-resolution taxonomic identification, understand the genetic diversity of CF-associated MAC isolates, and identify genetically similar strains among persons with CF for epidemiologic follow-up.

### Materials and Methods

Ethics approval for this work was obtained from the National Jewish Health Institutional Review Board (approval no. HS-3149). As part of Colorado Research and Development Program (<https://www.nationaljewishhealth.org/cocfrdp>), NTM isolates from US CFCCs were processed and biobanked with the goal of surveillance for genetically similar strains (Table 1). We cultured bacterial samples on Middlebrook 7H11 agar plates (ThermoFisher Scientific, <https://www.thermofisher.com>) supplemented with 10% oleic acid, albumin, dextrose, catalase growth supplement before subculturing single-colony isolates into Middlebrook 7H9 broth (ThermoFisher Scientific) supplemented with 10% albumin, dextrose, catalase growth supplement and 0.05% Tween 80 (Sigma-Aldrich, <https://www.sigmaaldrich.com>). We divided these cultures into 1-mL biobanked glycerol stock aliquot replicates that we stored at -20°C.

### DNA Extraction and Whole-Genome Sequencing

We extracted NTM DNA as described previously (19). We used NexteraXT DNA or DNA FLEX sample preparation (Illumina, <https://www.illumina.com>) to prepare WGS libraries and sequenced the libraries by using the Illumina MiSeq or HiSeq 2500. WGS data are available at the National Center for Biotechnology Information (BioProject no. PRJNA319839).

### Non-CF Sample Acquisition

To place RDP isolates in context with zoonotic, environmental, and clinical samples from around the world, we included additional MAC isolates with existing WGS in the study. We downloaded 874 MAC genomes from the National Center for Biotechnology Information, including MAV (559 total; 42 environmental, 467 non-CF clinical, and 50 zoonotic),

MCHIM (114 total; 3 environmental and 111 non-CF clinical), and MINT (201 total; 4 environmental, 192 non-CF clinical, and 5 zoonotic) from 32 published studies (Appendix 1 Table 1, <https://wwwnc.cdc.gov/EID/article/27/11/21-0124-App1.xlsx>) for subsequent comparisons.

### MAC Species Identification

We trimmed sequence reads of adapters and base calls with quality scores <Q20 by using Skewer (20). We then assembled trimmed reads into scaffolds by using Unicycler (21). We compared genome assemblies against a collection of reference genomes (Appendix 1 Table 1) to estimate average nucleotide identity (ANI) and assign a species call to each isolate (22,23). A cut-off ANI of ≥95% indicated the isolate and reference genome belonged to the same species.

### Phylogenomic Analysis

On the basis of taxonomic assignment, with the highest ANI score >95% for each genome, we mapped trimmed sequence reads to respective reference genomes (e.g., *M. avium* strain H87 [24]; *M. intracellulare* subsp. *chimaera* CDC 2015-22-71 [25]) by using Bowtie2 (26). We identified single-nucleotide polymorphisms (SNPs) as previously described (27).

By using the genome coordinates that correspond to the partial *rpoB* region used in clinical diagnostics, we extracted sequences from each MAC isolate. We compared the partial *rpoB* sequences from MAV, MCHIM, and MINT phylogenetically by using neighbor-joining and 250 bootstraps of the observed SNPs in MEGA (28).

To evaluate relationships between MAV from US CFCCs and global strains, we assessed the phylogenetic relationships to publicly available genomes from 559 non-CF MAV isolates, including 465 clinical, 42 environmental, and 50 zoonotic isolates from Japan, Germany, Belgium, the United Kingdom, the United States, and 12 other countries (Appendix 1 Table 1). To evaluate relationships between MCHIM from US CFCCs with US and global strains, we assessed the phylogenetic relationships to publicly available genomes from 114 non-CF MCHIM isolates, including 109 clinical and 5 environmental isolates from the United Kingdom, the United States,

**Table 1.** Number of MAC isolates in a study of MAC clusters in cystic fibrosis centers, United States\*

Category	MAV	MCHIM	MINT	Total
Patients with 1 isolate, no.	63	33	43	137
Patients with ≥2 isolates, no.	30	5	23	55
Total patients, no.	93	38	66	186
Total isolates, no.	186	44	134	364

\*Some patients have isolates from multiple MAC species. MAC, *Mycobacterium avium* complex; MAV, *M. avium*; MCHIM, *M. intracellulare* subsp. *chimaera*; MINT, *M. intracellulare* subsp. *intracellulare*.

Switzerland, South Korea, Canada, and South Africa (Appendix 1 Table 1). To evaluate relationships between MINT from US CFCCs with US and global strains, we assessed the phylogenetic relationships to publicly available genomes from 201 non-CF MINT isolates, including 192 clinical, 4 environmental, and 5 zoonotic isolates from China, the United Kingdom, South Korea, and the United States (Appendix 1 Table 1).

### Identifying Genetically Similar Isolate Clusters

To identify a SNP threshold for genetically similar isolates, we examined genomewide SNP distances between pairs of longitudinal isolates from the same person (within-patient isolates) and isolates from different persons (between-patient isolates) in the US CFCC MAC dataset, analogous to methods used previously for *M. abscessus* and MAV (29–33). The US CFCC MAC dataset included 56 persons with CF who had  $\geq 2$  isolates of the same species: 31 who had  $\geq 2$  MAV isolates, 5 who had  $\geq 2$  MCHIM isolates, and 23 who had  $\geq 2$  MINT isolates. We computed statistical comparisons between MAC groups by using Kruskal–Wallis tests. By using the distributions of within-patient and between-patient genomic SNPs (Figure 1, panel A), we defined a distance of  $\leq 20$  SNPs as the threshold difference for strain definition. We defined isolates found within a patient with a pairwise distance of  $>20$  SNPs as different strains. We notified CFCCs of genetically similar isolates and offered participation in site-specific epidemiologic investigations as part of the ongoing HALT-NTM trial (<https://clinicaltrials.gov/ct2/show/NCT04024423>) (34).

## Results

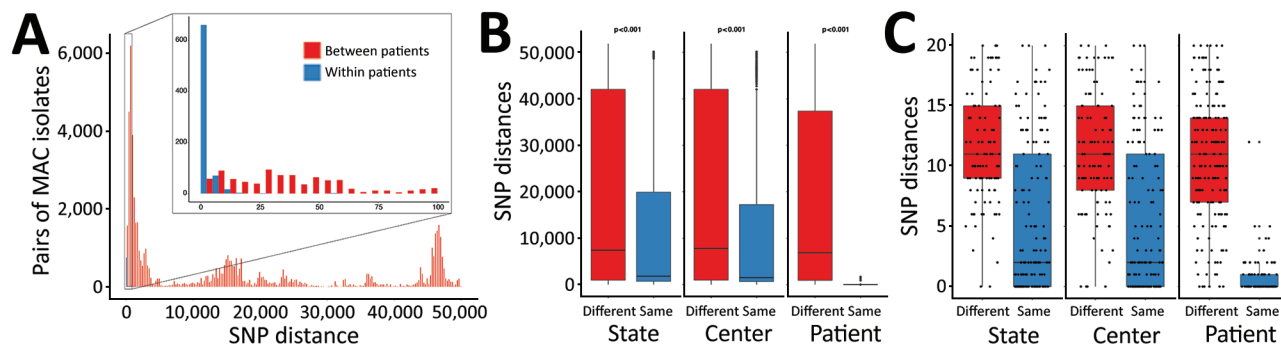
### Distribution of MAC Species in US Cystic Fibrosis Care Centers

We sequenced the genomes of 364 MAC isolates, including 186 MAV (51%), 134 MINT (37%), and 44 MCHIM (12%) (Table 1). More than half (101/186 [54%]) of persons with CF were women or girls (average age 35 years [range 9–88 years]). Isolates were analyzed from a total of 42 CFCCs and 22 states (Figure 2). Two-thirds (129/186 [69%]) of persons with CF had only 1 isolate sequenced, 21 had 2 isolates (21/186 [12%]), and 36 had  $\geq 3$  isolates (36/186 [19%]); collection dates spanned a range of 0 to 1,376 days between the first and last isolate collected (Figure 3). Most (132/186 [71%]) persons with samples analyzed were from 41 CFCCs in 21 states, and the remainder received care at 1 CFCC.

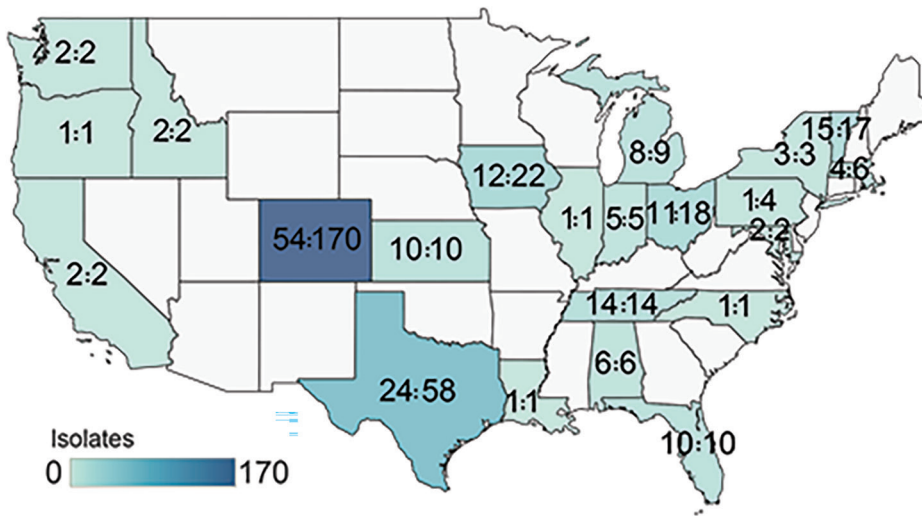
To evaluate taxonomic relationships of closely related taxa, we analyzed isolates from persons with CF, reference genomes for MAV (Figure 4, panel A; Appendix 1 Table 2), and type strains of MINT and MCHIM (Figure 4, panel B; Appendix 1 Table 1). The MAV phylogeny shows that most isolates from persons with CF are *M. avium* subsp. *hominissuis*, except for 1 isolate that was *M. avium* subsp. *avium* (Figure 4, panel A). The *M. intracellulare* phylogeny supports the taxonomy of 2 *M. intracellulare* subspecies, including MCHIM that is distinct from MINT (Figure 4, panel B).

### Polyclonal MAC infections in Persons with Cystic Fibrosis

Among 55 persons with CF who had  $\geq 2$  MAC isolates, we identified 15 (15/55 [27%]) who had multiple strains



**Figure 1.** Cluster analysis of MAC in persons with cystic fibrosis to identify recent shared ancestry in a study of MAC clusters in cystic fibrosis centers, United States. A) Pairwise SNP distances of *Mycobacterium avium* and *M. intracellulare* subsp. *chimaera*, and *M. intracellulare* subsp. *intracellulare* isolates from within same patients (blue) and between different patients (red). B) Pairwise SNP distances of all CFCC MAC by state, CFCC, and patient comparisons. Kruskal–Wallis rank-sum test p values for comparing mean differences between categories are specified above each comparison. C) Pairwise SNP distances of CFCC MAC by state, CFCC, and patient comparisons under the clustering threshold. Box and scatterplots in panels B and C show SNPs between isolates at the same versus different states, same versus different CFCC, and same versus different patients. Horizontal lines within boxes indicate medians; top and bottom of boxes indicate 25th and 75th percentiles; error bars indicate the maximum and minimum values observed in the distribution. CFCC, cystic fibrosis care center; MAC, *Mycobacterium avium* complex; SNP, single-nucleotide polymorphism.



**Figure 2.** Geographic distribution of 364 *Mycobacterium avium* complex isolates from 186 patients, by cystic fibrosis care center state of origin in study of *M. avium* complex clusters in cystic fibrosis centers, United States. Numbers in each state are the number of patients with cystic fibrosis and total isolates contributed from centers within the state.

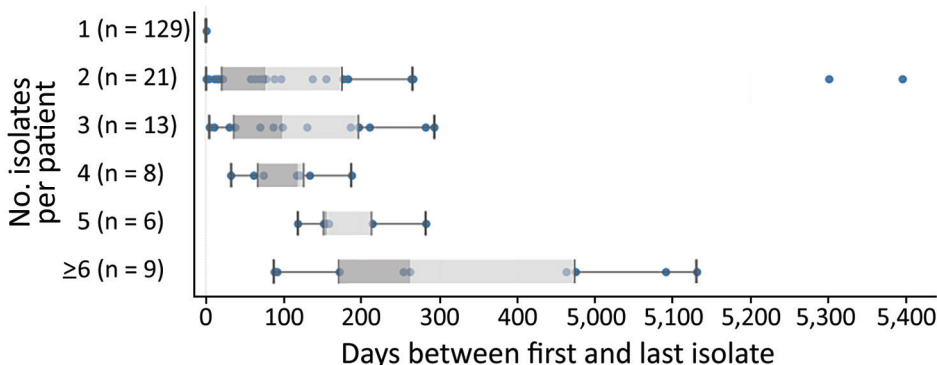
or species (Figure 5). Nine persons with CF (9/55 [16%]) had isolates from  $\geq 2$  MAC species; 1 (1/55 [2%]) had isolates of MAV, MCHIM, and MINT. Thirteen persons with CF who had MAV (13/30 [43%]) had  $\geq 2$  distinct MAV strains ( $>500$  SNPs apart). Among these 13 persons with CF, we observed an average of 2.3 (range 2–5) different strains/patient; average within-patient diversity was 3,384 SNPs. Two (9%) of 23 persons with CF had 2 different strains of MINT; no persons with MCHIM had multiple strains. In total, 15/55 persons with CF had  $>1$  strain or species, compared with 40/55 (73%) who had the same MAC strain isolated over time.

For the 15 persons with CF who had multiple strains of 1 MAC species, we generated time series plots of longitudinal isolates to visualize changes in strains over time (Figure 5; Appendix 2 Figure 1, <https://wwwnc.cdc.gov/EID/article/27/11/21-0124-App2.pdf>). The average time from first to last isolate collected was 259 days (range 5–1,262 days). In the case of the shortest interval, patient CF00193 was culture-positive with 2 different strains of MAV collected only 5 days apart. Patients CF00052, CF00060, and CF00193 each had 2 different strains of MAV collected

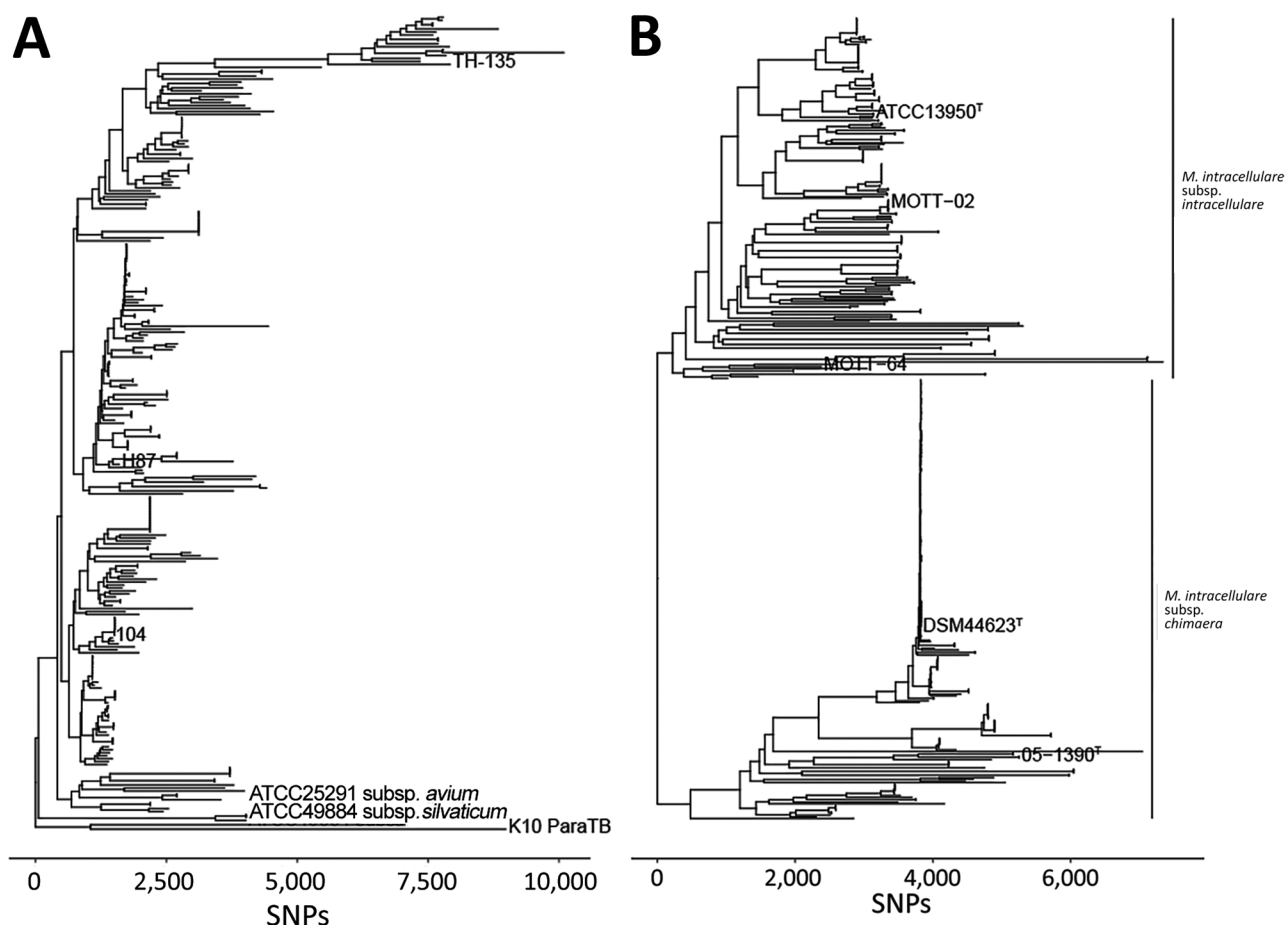
within 30-day windows. Patient CF00002 was culture-positive for 3 different strains of MAV within a single week and had 5 different strains over nearly 3.5 years. In the 2 persons with CF harboring multiple strains of MINT, the second strain was detected 42 days (CF00004) and 138 days (CF00131) after the first isolate collected. Patients CF00029 and CF00776 showed alternating strains over time, suggesting persistent mixed populations of MAV in the airway.

### Assessing Potential Transmission between Persons with Cystic Fibrosis

To evaluate routine molecular surveillance available in most diagnostic laboratories and compare it to the resolution afforded by WGS, we compared the *rpoB* partial sequences of each MAV, MCHIM, and MINT from US persons with CF. For MAV, 100% of patients belonged to 1 of 4 clusters (Appendix 2 Figure 2, panel A) based on analyses using *rpoB*, whereas 97.2% of MCHIM and 95.5% of MINT belonged to 5 clusters (Appendix 2 Figure 2, panel B). This result emphasizes that single-gene amplicon surveillance does not provide the resolution needed for genetic



**Figure 3.** Numbers of isolates per patient and days between the patient's first and last isolate collected in the isolate cohort in a study of *Mycobacterium avium* complex clusters in cystic fibrosis centers, United States. Vertical lines within boxes indicate medians; top and bottom of boxes indicate 25th and 75th percentiles; error bars indicate the maximum and minimum values observed in the distribution.



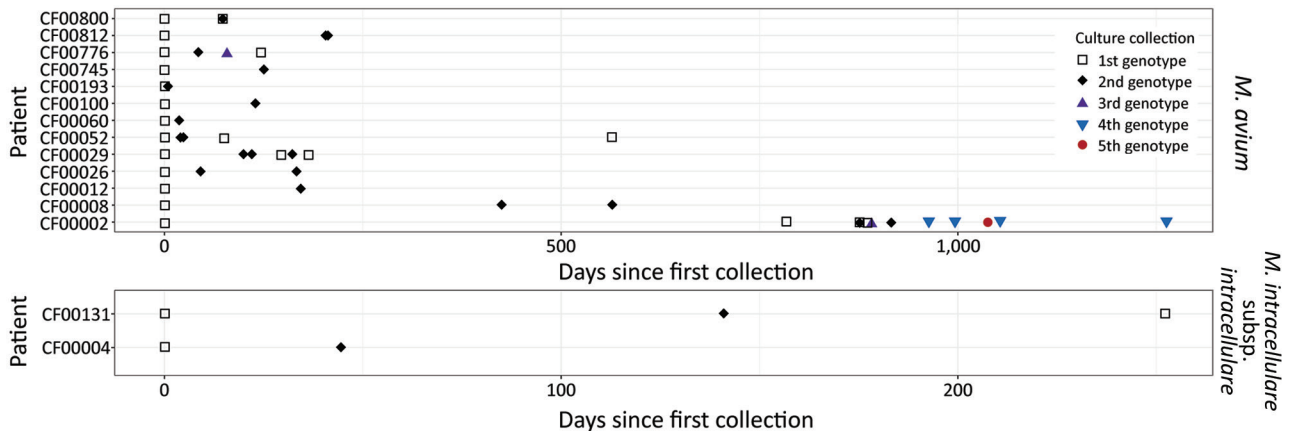
**Figure 4.** Phylogenetic relationships in a study of *Mycobacterium avium* complex isolates in cystic fibrosis care centers, United States. A) Phylogenetic tree of 207 *M. avium* isolates showing the relationships between *M. avium* cystic fibrosis care center and select non-cystic fibrosis, environmental, and zoonotic isolates. B). Phylogenetic tree of 235 isolates showing the relationships between cystic fibrosis care center and select non-cystic fibrosis *M. intracellulare* subsp. *chimaera* and *M. intracellulare* subsp. *intracellulare* isolates. Former species *M. yongonense* type strain 05-1380<sup>T</sup> was also included as part of *M. intracellulare* subsp. *chimaera* to reflect current taxonomy. SNP, single-nucleotide polymorphism.

surveillance of NTM MAC species, whereas WGS does provide the necessary resolution.

To examine potential transmission of MAC isolates between persons with CF, we identified 20 SNPs as the threshold for recent shared ancestry on the basis of the distribution of SNPs among longitudinal isolates collected over time (Figure 1, panel A). By using this threshold, we identified a total of 18 genetically similar clusters, including 3 MAV, 5 MCHIM, and 10 MINT clusters (Figure 6). Of the 3 MAV clusters, 2 clusters consisting of 6 patients receiving treatment at 1 CFCC, and a third cluster consisting of 2 patients from a second CFCC. Most patients (15/27 [56%]) in 3/5 MCHIM clusters received treatment in the same CFCCs, whereas the remaining isolates in clusters originated from patients attending different CFCCs. Alternatively, a minority (4/21 [19%]) of patients in 2/10 MINT clusters

received treatment in the same CFCC, suggesting that MINT may have different transmission routes compared with MAV or MCHIM. Among the entire US CFCC MAC dataset, 8/93 persons with MAV (9%), 15/36 with MCHIM (42%), and 4/66 with MINT (6%) belonged to clusters within the threshold of 20 SNPs and were treated at the same CFCCs, triggering epidemiologic follow-up in the HALT-NTM Trial (34). By using a 10-SNP threshold, we identified 2 *M. avium* clusters, 5 *M. chimaera*, and 6 *M. intracellulare* clusters (Appendix Figure 7). Overall, 4 patients included in 2 clusters defined by a 20-SNP threshold are removed when the threshold is reduced to 10 SNPs.

Overall, 27/186 persons with CF (15%) had MAC isolates that were genetically similar and received treatment at the same CFCC. Isolates collected within the same center were more similar than isolates



**Figure 5.** Polyclonal *Mycobacterium avium* complex (MAC) infections in 15 persons with CF in a study of MAC clusters in CF centers, United States. Persons with CF who had >1 MAC isolate were analyzed for the presence of multiple strains within a given MAC species. For *M. avium* (top) and *M. intracellulare* subsp. *intracellulare* (bottom), each row on the y-axis is a person with CF, and the x-axis represents the number of days after the first MAC isolate with whole-genome sequencing was collected. Each point represents a sequenced isolate and the shape represents a unique genotype. The plots do not represent all positive cultures in the patients' histories, but they illustrate how strains change, alternate, or both over time. In some cases, different strains were isolated on the same day or within a 1-week period. CF, cystic fibrosis.

collected from the same state ( $p = 0.014$ ), whereas the mean SNPs observed between isolates coming from different centers were not significantly different from those coming from different states (Figure 1, panel B). The mean SNP differences observed between nearest-neighbor clustered MAC isolates from the same versus different CFCCs (5.47 vs. 11.21 SNPs;  $p < 0.001$ ) and the same versus different states (5.45 vs. 11.46 SNPs;  $p < 0.001$ ) were both significant (Figure 1, panel C). Only 2 clustered patient pairs (4/186 [2%]) were identified between different centers within a state, suggesting that clustering is more localized to CFCCs than to states.

For isolate clusters that included  $\geq 3$  isolates, we visualized the isolate relationships as phylogenetic clades (Figure 7). The patient with the isolate nearest to the base of each clade is ancestral to all descendants, and therefore is a potential source of transmission between the subsequent patients in the cluster. For example, patient CF00002 was the potential source of 2 separate clusters of MAV and MCHIM. In the MAV cluster (Figure 7, panel A), 4 isolates from patient CF00002 were ancestral to isolates from 3 other patients (CF00231, CF00776, and CF00812). In the MCHIM cluster, 2 isolates from patient CF00002 were ancestral to 1 isolate from patient CF00966 (Figure 7, panel E). Ancestral isolates and hypotheses about the order in which transmission events occurred can similarly be deduced for an additional MAV cluster (Figure 7, panel B), 3 MCHIM clusters (Figure 7, panels C–E), and 1 MINT cluster (Figure 7, panel F).

## MAV

By using a genetic similarity threshold of 20 SNPs, we observed limited instances of genetic similarity between US CFCC MAV isolates from 11 persons with CF and 21 non-CF isolates (Appendix Figure 3). Four persons with CF had genetically similar MAV isolates to an environmental isolate collected from a household dust sample in Germany (Table 2; Appendix 2 Figure 4). Comparisons of US persons with CF MAV isolates to non-US clinical and zoonotic MAV isolates revealed similarities with 17 clinical isolates from patients in 6 countries (Belarus, Canada, Germany, Norway, United Kingdom, and United States), 3 zoonotic isolates from 2 birds (35), and 1 from an elephant. Overall, only 11/93 (12%) of persons with CF shared genetically similar isolates with non-CF MAV isolates.

## MCHIM

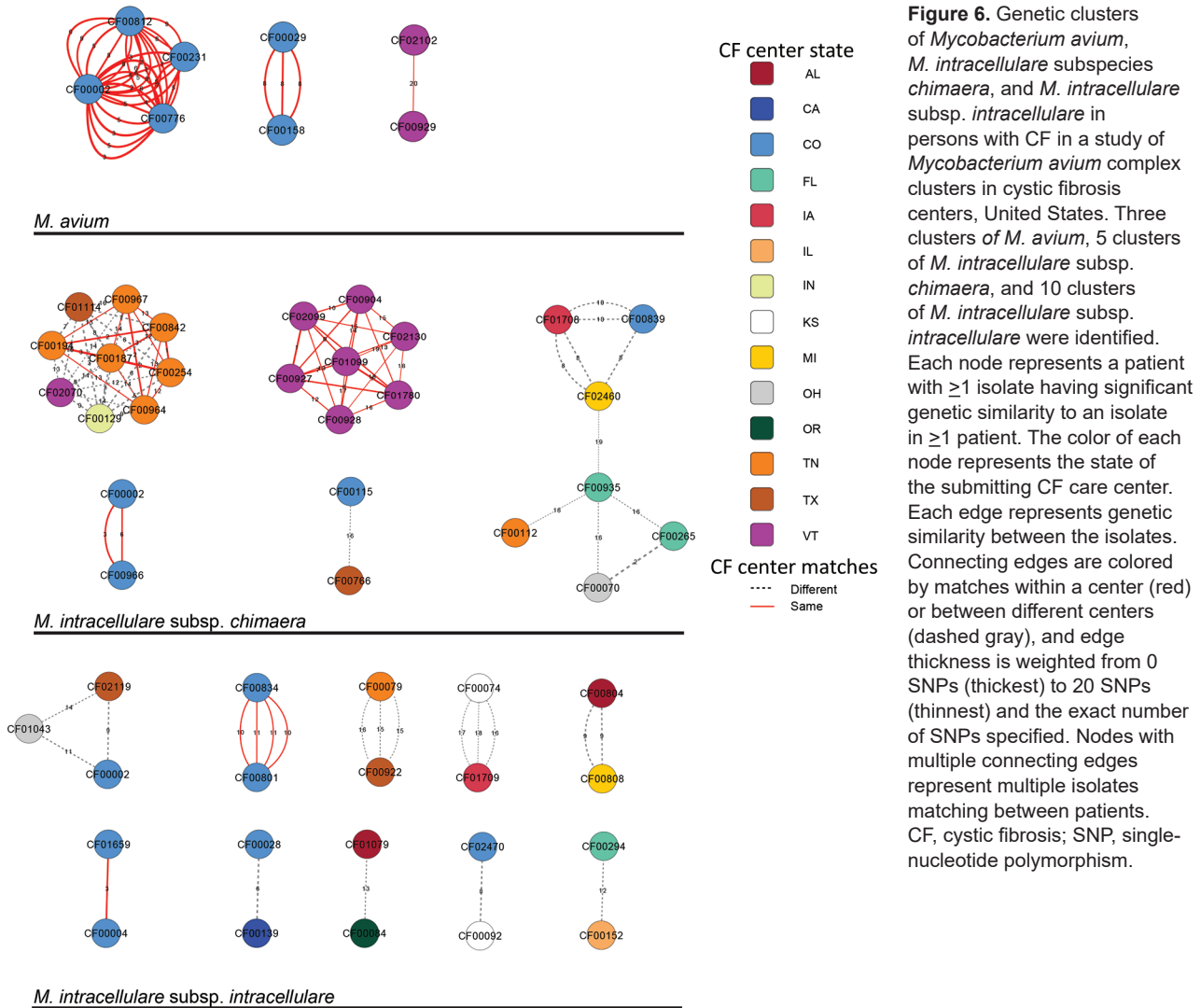
A total of 30 MCHIM isolates from 28 persons with CF were similar to 37 non-CF isolates (Appendix 2 Figure 3). Matches to US CFCC isolates also include the MCHIM type strain DSM44623<sup>T</sup>, 21 isolates from Oxford Hospital (Oxford, UK), and isolates from patients treated in Canada, Hawaii, and Virginia (Table 2; Appendix 2 Figure 5). US CFCC MCHIM isolates were all genetically different from isolates derived from contaminated heater-cooler units (36). No other environmental MCHIM isolates were available for comparisons. In total, 28/38 (74%) persons with CF and MCHIM had genetically similar isolates to non-CF isolates.

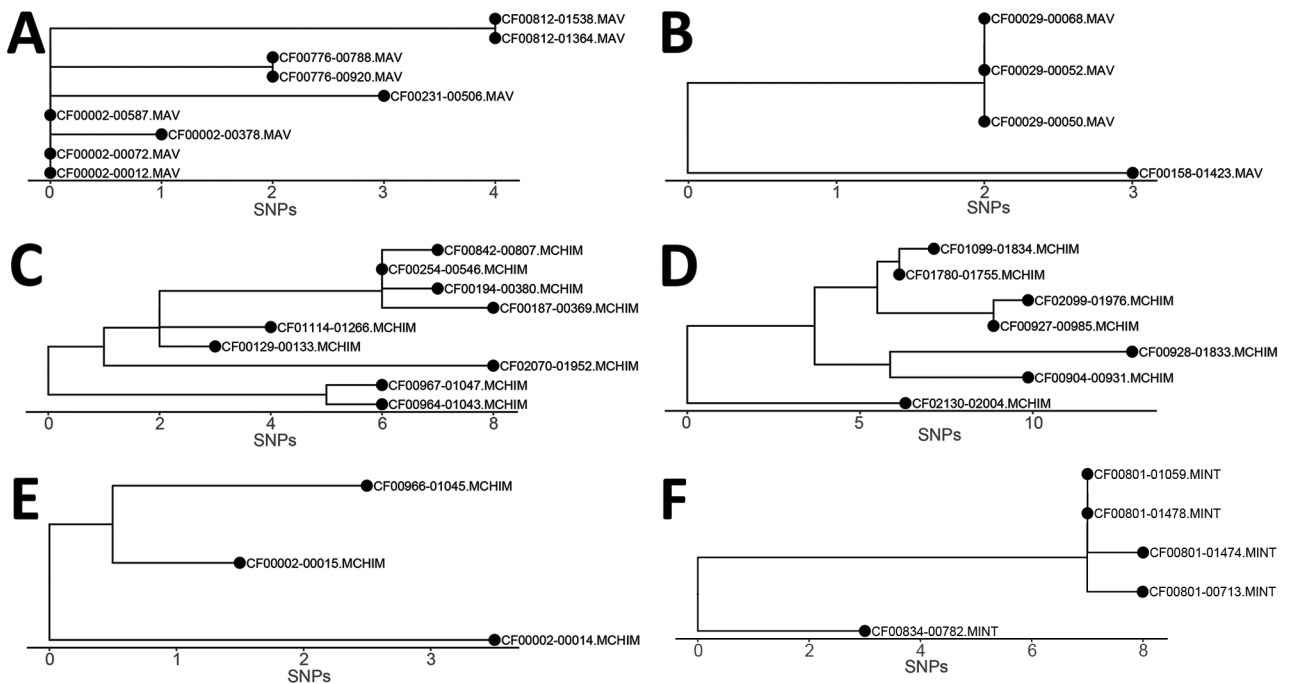
MINT

For MINT, we observed genetic similarities between isolates from 14 persons with CF and 24 non-CF isolates from North America, Europe, and Asia (Appendix 2 Figure 3). Eight MINT isolates were genetically similar to reference isolates, including MINT MOTT-02 (37), NCTC-13025 (38), and 22 nonpatient isolates from Michigan, Virginia, South Korea, and the United Kingdom (Table 2; Appendix 2 Figure 5). We did not observe similarities between environmental MINT and US CFCC isolates. Comparisons of US CFCC MINT isolates with zoonotic isolates identified similarity with isolates collected from a bird in a California zoo and the other from a penguin in a New York State zoo (35,39). Overall, 14/66 (21%) persons with CF and MINT had isolates with genetically similar matches to our non-CF isolate sample set.

Discussion

This study provides evidence of highly similar MAC isolates among persons with CF. However, the isolates from most MAC infections appear to be independently acquired and unclustered. We identified 18 genetically similar isolate clusters involving 54 persons with CF (including 8 patients with MAV, 27 patients with MCHIM, and 21 patients with MINT) within our threshold of recent shared ancestry ( $\leq 20$  SNPs). We further determined that 8 of the identified clusters (8/18 [44%]) included 26 patients that received treatment at the same CFCCs. Person-to-person transmission may have occurred among those persons, and the genetic clusters are undergoing epidemiologic investigation (34). Epidemiologic follow-up will help us understand if genetic similarity is related to acquisition through common geography and environments. Most persons with CF (160/186 [86%]) in our study





**Figure 7.** Phylogenetic visualization of *Mycobacterium avium* complex clusters in persons with CF in a study of *M. avium* complex clusters in CF centers, United States. Clusters with  $\geq 3$  isolates were visualized as clades to show the transition of patients' isolates over time. A) Cluster of 4 persons with MAV. B) Cluster of 2 persons with MAV. C) Cluster of 9 persons with MCHIM. D) Cluster of 7 persons with MCHIM. E) Cluster of 2 persons with MCHIM. F) Cluster of 2 persons with MINT. CF, cystic fibrosis; MAV, *M. avium*; MCHIM, *M. intracellulare* subsp. *chimaera*; MINT, *M. intracellulare* subsp. *intracellulare*; SNP, single-nucleotide polymorphism.

did not share similar strains; thus, we infer that most persons with CF do not transmit strains person-to-person or share acquisition sources of MAC.

In contrast with the clonality observed in *M. abscessus* (27,29), 27% of patients with MAC cultured multiple strains over time, as has also been observed for *Staphylococcus aureus* infections in persons with CF (40). This observation was considerably lower than the proportion of polyclonal MAC infections previously observed in patients with non-CF NTM lung disease (29). Although the analysis of single isolates instead of colony sweeps provides the clarity to genetically identify transmission clusters, it may underestimate the diversity of MAC populations present in patient airways. We surmise that MAV isolates found in most US persons with CF

probably derive from the independent acquisition (or acquisitions) of strains in the environment. This interpretation is consistent with previously observed instances of genetically matched environmental and patient MAV isolates (10,13,14,17,30,41); however, it does not exclude the hypothesis of person-to-person transmission in persons with CF. Two hypotheses can explain the observations of multiple genotypes and species in persons with CF: patients were originally infected with multiple genotypes of MAC that were selected for during infection and treatment, or patients cleared the original infection and subsequently acquired a new, independent genotype. Our analyses provide evidence for both scenarios (Figure 5), though with limited sample sizes. Further studies of within-patient population diversity with

**Table 2.** Persons with CF with genetically similar MAC isolates compared with publicly available non-CF isolates collected from environmental, clinical, and zoonotic sources in a study of MAC clusters in CF centers, United States\*

Taxon	No. (%)			Total similar CF patients
	Environmental	Non-CF clinical	Zoonotic	
MAV	4/93 (5)	9/93 (12)	3/93 (3)	11/93 (12)
MCHIM	0/38 (21)	28/38 (61)	0/38 (0)	28/38 (74)
MINT	0/66 (0)	14/66 (24)	3/66 (5)	14/66 (21)
Total				52/186 (28)

\*Some patients have isolates that match to an isolate (or isolates) in  $>1$  category. CF, cystic fibrosis; MAC, *Mycobacterium avium* complex; MAV, *M. avium*; MCHIM, *M. intracellulare* subsp. *chimaera*; MINT, *M. intracellulare* subsp. *intracellulare*.

corresponding environmental sampling are needed to address these questions.

Our WGS analysis of 364 MAC isolates, sent from 42 CFCCs in 23 states across the United States as part of a voluntary nationwide surveillance program, enabled us to examine genetic relationships among US isolates. WGS analyses greatly reduced the sizes of MAC clusters identified in US persons with CF compared with *rpoB* sequence information alone, highlighting the value of WGS resolution for epidemiologic follow-up. We also compared CF MAC isolates to isolates from previous studies, including those from environmental, zoonotic, and non-CF clinical sources. In our study, US MAV isolates from persons with CF were mostly distinct from non-CF clinical, environmental, and zoonotic samples from the United States (30), Europe (42–44), and Asia (12,37,45,46), although 12% of patients in our study had genetic matches to non-CF isolates. This finding is consistent with observations of human patients and animals harboring identical MAV in Europe (8,9,12,13,42,47,48). Similarly, only 21% of persons with CF and MINT had genetically similar isolates to non-CF samples, primarily clinical isolates. Few publicly available environmental isolates of MINT were available for comparison because of the lack of MINT found in water sources (49), suggesting that persons with CF likely acquire their MAV and MINT infections from nonhuman reservoirs that were not identified in this study.

In contrast, we observed many matches of MCHIM between CF and non-CF isolates. Indeed, a high proportion of MCHIM from US persons with CF (74% of patients) had matches to non-CF clinical isolates relative to MAV or MINT. One hypothesis to explain clustering of MCHIM is that the observed strains are well-adapted to colonize and persist in a human host. Alternatively, the high genetic similarity of MCHIM isolates may also suggest a lineage that has recently come to prominence in North America. Additional environmental and zoonotic sampling of MAV, MCHIM, and MINT isolates in the United States will be needed to better understand the species-specific risks of MAC infection from these sources.

Our study has some limitations. First, our empirically defined SNP threshold for recent common ancestry is specific for our patient cohort and is limited by the number of persons with CF with  $\geq 2$  isolates and the duration of sampling time frames. Thus, our threshold may miss transmission events that occurred before the sampling period. Second, despite observing genetic matches, epidemiologic links are required to support transmission. Our epidemiologic

data were limited to isolate collection date and the CFCC where patients received care. Therefore, our analyses provide hypotheses for traditional epidemiologic follow-up at CFCCs that was beyond the scope of our current project but is being addressed in the HALT-NTM Trial (34). Third, the publicly available datasets did not allow a uniform comparison to non-CF clinical, environmental, or zoonotic isolates from each CFCC region for each species.

Our research study discovered potential instances of transmission between patients and assessed the dynamics of MAC infections in persons with CF. The findings of our US-based surveillance work in persons with CF were not possible without the resolution of WGS and underscore the need for continued epidemiologic follow-up in patients with MAC lung disease, with and without CF, to assist infectious disease control measures and limit the spread of MAC infections where possible.

This work was supported by the US Cystic Fibrosis Foundation's Colorado Research and Development Program (grant no. NICK15RO) and the Cystic Fibrosis National Resource Centers (grant nos. NICK20Y2SVC and NICK20Y2OUT). R.M.D was funded by National Institutes of Health–National Institute of Allergy and Infectious Diseases grant no. K01-AI125726. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## About the Author

Dr. Hasan is a bioinformatics analyst at the Center for Genes, Environment, and Health at National Jewish Health, Denver, Colorado. His main research interests are population genomics, acquisition, and environmental sources of nontuberculous mycobacteria.

## References

1. Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, et al.; ATS Mycobacterial Diseases Subcommittee; American Thoracic Society; Infectious Disease Society of America. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am J Respir Crit Care Med*. 2007; 175:367–416. <https://doi.org/10.1164/rccm.200604-571ST>
2. Adjemian J, Olivier KN, Seitz AE, Holland SM, Prevots DR. Prevalence of nontuberculous mycobacterial lung disease in U.S. Medicare beneficiaries. *Am J Respir Crit Care Med*. 2012;185:881–6. <https://doi.org/10.1164/rccm.201111-2016OC>
3. Castejon M, Menéndez MC, Comas I, Vicente A, Garcia MJ. Whole-genome sequence analysis of the *Mycobacterium avium* complex and proposal of the transfer of *Mycobacterium yongonense* to *Mycobacterium intracellulare* subsp. *yongonense* subsp. nov. *Int J Syst Evol Microbiol*. 2018;68:1998–2005. <https://doi.org/10.1099/ijsem.0.002767>

4. Nouioui I, Carro L, García-López M, Meier-Kolthoff JP, Woyke T, Kyrpides NC, et al. Genome-based taxonomic classification of the phylum *Actinobacteria*. *Front Microbiol*. 2018;9:2007. <https://doi.org/10.3389/fmicb.2018.02007>
5. Tortoli E, Meehan CJ, Grottola A, Fregni Serpini G, Fabio A, Trovato A, et al. Genome-based taxonomic revision detects a number of synonymous taxa in the genus *Mycobacterium*. *Infect Genet Evol*. 2019;75:103983. <https://doi.org/10.1016/j.meegid.2019.103983>
6. van Ingen J, Turenne CY, Tortoli E, Wallace RJ Jr, Brown-Elliott BA. A definition of the *Mycobacterium avium* complex for taxonomical and clinical purposes, a review. *Int J Syst Evol Microbiol*. 2018;68:3666–77. <https://doi.org/10.1099/ijsem.0.003026>
7. Adjemian J, Olivier KN, Prevots DR. Epidemiology of pulmonary nontuberculous mycobacterial sputum positivity in patients with cystic fibrosis in the United States, 2010–2014. *Ann Am Thorac Soc*. 2018;15:817–26. <https://doi.org/10.1513/AnnalsATS.201709-727OC>
8. Arikawa K, Ichijo T, Nakajima S, Nishiuchi Y, Yano H, Tamaru A, et al. Genetic relatedness of *Mycobacterium avium* subsp. *hominissuis* isolates from bathrooms of healthy volunteers, rivers, and soils in Japan with human clinical isolates from different geographical areas. *Infect Genet Evol*. 2019;74:103923. <https://doi.org/10.1016/j.meegid.2019.103923>
9. Iwamoto T, Nakajima C, Nishiuchi Y, Kato T, Yoshida S, Nakanishi N, et al. Genetic diversity of *Mycobacterium avium* subsp. *hominissuis* strains isolated from humans, pigs, and human living environment. *Infect Genet Evol*. 2012;12:846–52. <https://doi.org/10.1016/j.meegid.2011.06.018>
10. Nishiuchi Y, Iwamoto T, Maruyama F. Infection sources of a common non-tuberculous mycobacterial pathogen, *Mycobacterium avium* complex. *Front Med (Lausanne)*. 2017;4:27. <https://doi.org/10.3389/fmed.2017.00027>
11. Tzou CL, Dirac MA, Becker AL, Beck NK, Weigel KM, Meschke JS, et al. Association between *Mycobacterium avium* complex pulmonary disease and mycobacteria in home water and soil. *Ann Am Thorac Soc*. 2020;17:57–62. <https://doi.org/10.1513/AnnalsATS.201812-915OC>
12. Yano H, Iwamoto T, Nishiuchi Y, Nakajima C, Starkova DA, Mokrousov I, et al. Population structure and local adaptation of MAC lung disease agent *Mycobacterium avium* subsp. *hominissuis*. *Genome Biol Evol*. 2017;9:2403–17. <https://doi.org/10.1093/gbe/evx183>
13. Nishiuchi Y, Maekura R, Kitada S, Tamaru A, Taguri T, Kira Y, et al. The recovery of *Mycobacterium avium-intracellulare* complex (MAC) from the residential bathrooms of patients with pulmonary MAC. *Clin Infect Dis*. 2007;45:347–51. <https://doi.org/10.1086/519383>
14. Falkinham JOI III. Nontuberculous mycobacteria from household plumbing of patients with nontuberculous mycobacteria disease. *Emerg Infect Dis*. 2011;17:419–24. <https://doi.org/10.3201/eid1703.101510>
15. Feazel LM, Baumgartner LK, Peterson KL, Frank DN, Harris JK, Pace NR. Opportunistic pathogens enriched in showerhead biofilms. *Proc Natl Acad Sci U S A*. 2009;106:16393–9. <https://doi.org/10.1073/pnas.0908446106>
16. Thomson R, Tolson C, Carter R, Coulter C, Huygens F, Hargreaves M. Isolation of nontuberculous mycobacteria (NTM) from household water and shower aerosols in patients with pulmonary disease caused by NTM. *J Clin Microbiol*. 2013;51:3006–11. <https://doi.org/10.1128/JCM.00899-13>
17. De Groote MA, Pace NR, Fulton K, Falkinham JOI III. Relationships between *Mycobacterium* isolates from patients with pulmonary mycobacterial infection and potting soils. *Appl Environ Microbiol*. 2006;72:7602–6. <https://doi.org/10.1128/AEM.00930-06>
18. Honda JR, Hasan NA, Davidson RM, Williams MD, Epperson LE, Reynolds PR, et al. Environmental nontuberculous mycobacteria in the Hawaiian Islands. *PLoS Negl Trop Dis*. 2016;10:e0005068. <https://doi.org/10.1371/journal.pntd.0005068>
19. Epperson LE, Strong M. A scalable, efficient, and safe method to prepare high quality DNA from mycobacteria and other challenging cells. *J Clin Tuberc Other Mycobact Dis*. 2020;19:100150. <https://doi.org/10.1016/j.jctube.2020.100150>
20. Jiang H, Lei R, Ding S-W, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*. 2014;15:182. <https://doi.org/10.1186/1471-2105-15-182>
21. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol*. 2017;13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
22. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol*. 2007;57:81–91. <https://doi.org/10.1099/ijms.0.64483-0>
23. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A*. 2009;106:19126–31. <https://doi.org/10.1073/pnas.0906412106>
24. Zhao X, Epperson LE, Hasan NA, Honda JR, Chan ED, Strong M, et al. Complete genome sequence of *Mycobacterium avium* subsp. *hominissuis* strain H87 isolated from an indoor water sample. *Genome Announc*. 2017;5:e00189–17. <https://doi.org/10.1128/genomeA.00189-17>
25. Hasan NA, Lawsin A, Perry KA, Alyanak E, Toney NC, Malecha A, et al. Complete genome sequence of *Mycobacterium chimaera* strain CDC 2015–22–71. *Genome Announc*. 2017;5:e00693–17. <https://doi.org/10.1128/genomeA.00693-17>
26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9. <https://doi.org/10.1038/nmeth.1923>
27. Davidson RM, Hasan NA, Epperson LE, Benoit JB, Kammlade SM, Levin AR, et al. Population genomics of *Mycobacterium abscessus* from United States cystic fibrosis care centers. *Ann Am Thorac Soc*. 2021 Apr 15 [Epub ahead of print]. <https://doi.org/10.1513/AnnalsATS.202009-1214OC>
28. Stecher G, Tamura K, Kumar S. Molecular evolutionary genetics analysis (MEGA) for macOS. *Mol Biol Evol*. 2020;37:1237–9. <https://doi.org/10.1093/molbev/msz312>
29. Bryant JM, Brown KP, Burbau S, Everall I, Belardinelli JM, Rodriguez-Rincon D, et al. Stepwise pathogenic evolution of *Mycobacterium abscessus*. *Science*. 2021;372:eabb8699. <https://doi.org/10.1126/science.abb8699>
30. Lande L, Alexander DC, Wallace RJ Jr, Kwiat R, Iakhiaeva E, Williams M, et al. *Mycobacterium avium* in community and household water, suburban Philadelphia, Pennsylvania, USA, 2010–2012. *Emerg Infect Dis*. 2019;25:473–81. <https://doi.org/10.3201/eid2503.180336>
31. Bryant JM, Grogono DM, Greaves D, Foweraker J, Roddick I, Inns T, et al. Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study. *Lancet*. 2013;381:1551–60. [https://doi.org/10.1016/S0140-6736\(13\)60632-7](https://doi.org/10.1016/S0140-6736(13)60632-7)

32. Tortoli E, Kohl TA, Trovato A, Baldan R, Campana S, Cariani L, et al. *Mycobacterium abscessus* in patients with cystic fibrosis: low impact of inter-human transmission in Italy. *Eur Respir J*. 2017;50:1602525. <https://doi.org/10.1183/13993003.02525-2016>
33. Yoon J-K, Kim TS, Kim J-I, Yim J-J. Whole genome sequencing of nontuberculous mycobacterium (NTM) isolates from sputum specimens of co-habiting patients with NTM pulmonary disease and NTM isolates from their environment. *BMC Genomics*. 2020;21:322. <https://doi.org/10.1186/s12864-020-6738-2>
34. Gross JE, Martiniano SL, Nick JA. Prevention of transmission of *Mycobacterium abscessus* among patients with cystic fibrosis. *Curr Opin Pulm Med*. 2019;25:646–53. <https://doi.org/10.1097/MCP.0000000000000621>
35. Pfeiffer W, Braun J, Burchell J, Witte CL, Rideout BA. Whole-genome analysis of mycobacteria from birds at the San Diego Zoo. *PLoS One*. 2017;12:e0173464. <https://doi.org/10.1371/journal.pone.0173464>
36. Hasan NA, Epperson LE, Lawsin A, Rodger RR, Perkins KM, Halpin AL, et al. Genomic analysis of cardiac surgery-associated *Mycobacterium chimaera* infections, United States. *Emerg Infect Dis*. 2019;25:559–63. <https://doi.org/10.3201/eid2503.181282>
37. Kim B-J, Choi B-S, Lim J-S, Choi I-Y, Lee J-H, Chun J, et al. Complete genome sequence of *Mycobacterium intracellulare* clinical strain MOTT-02. *J Bacteriol*. 2012;194:2771. <https://doi.org/10.1128/JB.00365-12>
38. Chand M, Lamagni T, Kranzer K, Hedge J, Moore G, Parks S, et al. Insidious risk of severe *Mycobacterium chimaera* infection in cardiac surgery patients. *Clin Infect Dis*. 2017;64:335–42. <https://doi.org/10.1093/cid/ciw754>
39. Rivas AE, Hollinger C, Oehler DA, Robbe-Austerman S, Paré JA. Diagnosis and management of mycobacteriosis in a colony of little penguins (*Eudyptula minor*). *J Zoo Wildl Med*. 2019;50:427–36. <https://doi.org/10.1638/2018-0190>
40. Long DR, Wolter DJ, Lee M, Precit M, McLean K, Holmes E, et al. Polyclonality, shared strains, and convergent evolution in chronic CF *S. aureus* airway infection. *Am J Respir Crit Care Med*. 2021;203:1127–37. <https://doi.org/10.1164/rccm.202003-0735OC>
41. Feazel LM, Baumgartner LK, Peterson KL, Frank DN, Harris JK, Pace NR. Opportunistic pathogens enriched in showerhead biofilms. *Proc Natl Acad Sci U S A*. 2009;106:16393–9. <https://doi.org/10.1073/pnas.0908446106>
42. Bruffaerts N, Vluggen C, Duytschaever L, Mathys V, Saegerman C, Chapeira O, et al. Genome sequences of four strains of *Mycobacterium avium* subsp. *hominissuis*, isolated from swine and humans, differing in virulence in a murine intranasal infection model. *Genome Announc*. 2016;4:e00533–16. <https://doi.org/10.1128/genomeA.00533-16>
43. Lahiri A, Kneisel J, Kloster I, Kamal E, Lewin A. Abundance of *Mycobacterium avium* ssp. *hominissuis* in soil and dust in Germany – implications for the infection route. *Lett Appl Microbiol*. 2014;59:65–70. <https://doi.org/10.1111/lam.12243>
44. Sanchini A, Semmler T, Mao L, Kumar N, Dematheis F, Tandon K, et al. A hypervariable genomic island identified in clinical and environmental *Mycobacterium avium* subsp. *hominissuis* isolates from Germany. *Int J Med Microbiol*. 2016;306:495–503. <https://doi.org/10.1016/j.ijmm.2016.07.001>
45. Uchiya K, Takahashi H, Yagi T, Moriyama M, Inagaki T, Ichikawa K, et al. Comparative genome analysis of *Mycobacterium avium* revealed genetic diversity in strains that cause pulmonary and disseminated disease. *PLoS One*. 2013;8:e71831. <https://doi.org/10.1371/journal.pone.0071831>
46. Uchiya KI, Tomida S, Nakagawa T, Asahi S, Nikai T, Ogawa K. Comparative genome analyses of *Mycobacterium avium* reveal genomic features of its subspecies and strains that cause progression of pulmonary disease. *Sci Rep*. 2017;7:39750. <https://doi.org/10.1038/srep39750>
47. Ichikawa K, van Ingen J, Koh WJ, Wagner D, Salfinger M, Inagaki T, et al. Genetic diversity of clinical *Mycobacterium avium* subsp. *hominissuis* and *Mycobacterium intracellulare* isolates causing pulmonary diseases recovered from different geographical regions. *Infect Genet Evol*. 2015;36:250–5. <https://doi.org/10.1016/j.meegid.2015.09.029>
48. Vluggen C, Soetaert K, Duytschaever L, Denoël J, Fauville-Dufaux M, Smeets F, et al. Genotyping and strain distribution of *Mycobacterium avium* subspecies *hominissuis* isolated from humans and pigs in Belgium, 2011–2013. *Euro Surveill*. 2016;21:30111. <https://doi.org/10.2807/1560-7917.ES.2016.21.3.30111>
49. Wallace RJ Jr, Iakhiaeva E, Williams MD, Brown-Elliott BA, Vasireddy S, Vasireddy R, et al. Absence of *Mycobacterium intracellulare* and presence of *Mycobacterium chimaera* in household water and biofilm samples of patients in the United States with *Mycobacterium avium* complex respiratory disease. *J Clin Microbiol*. 2013;51:1747–52. <https://doi.org/10.1128/JCM.00186-13>

---

Address for correspondence: Nabeeh A. Hasan, Center for Genes, Environment, and Health, National Jewish Health, 1400 Jackson St, Denver, CO 80206, USA; email: hasann@njhealth.org

# Population Genomics and Inference of *Mycobacterium avium* Complex Clusters in Cystic Fibrosis Care Centers, United States

## Appendix 2

### Whole-Genome Sequence and SNP Phylogenetic Analysis

Illumina sequence reads were trimmed using Skewer and assembled into draft genomes using Unicycler (1). For each isolate, genome assemblies were compared against a collection of reference nontuberculous mycobacteria (NTM) genomes to assign taxonomic identification based upon the highest estimate the average nucleotide identity (ANI) above 95%.

Based upon the highest ANI score above 95% for each genome, the taxonomic assignment determined the reference genome each isolate's trimmed sequence reads was mapped to (e.g., *M. avium* isolates mapped to *M. avium* strain H87 or *M. intracellulare* subsp. *chimaera* and *M. intracellulare* subsp. *intracellulare* isolates were mapped to *M. intracellulare* subsp. *chimaera* CDC 2015–22–71).

Single nucleotide polymorphisms were called using SAMtools mpileup program (2) and were filtered using a custom perl script using the following parameters: SNP quality score of 20, minimum of 4x read depth, the majority of base calls support the variant base, and less than 25% of variant calls occur at the beginning or end of the sequence reads. A multi-fasta sequence alignment was created from concatenated base calls from all isolates. SNP-sites was used to filter positions in the genome in which at least one strain differed from the reference genome, and for which high quality variant and/or reference calls were present for all other isolates (3).

To evaluate relationships between MAV from US CF care centers (CFCCs) and global strains, we assessed the phylogenetic relationships to publicly available genomes from 559 non-CF MAV isolates, including 465 clinical, 42 environmental, and 50 zoonotic isolates from Japan

(4), Germany (5,6), Belgium (7,8), Great Britain (9), the U.S (10–13). and 12 other countries (Appendix 1 Table 2, <https://wwwnc.cdc.gov/EID/article/27/11/21-0124-App1.xlsx>).

To evaluate relationships between MCHIM from US CF Care Centers with US and global strains, we assessed the phylogenetic relationships to publicly available genomes from 114 non-CF MCHIM isolates, including 109 clinical, 5 environmental isolates, from the United Kingdom (14–20), the U.S (10,12,21), Switzerland (15,22,23), Korea (24), Canada (25), South Africa (unpublished; Appendix 1 Table 2).

To evaluate relationships between MINT from US CFCCs with US and global strains, we assessed the phylogenetic relationships to publicly available genomes from 201 non-CF MINT isolates, including 192 clinical, 4 environmental, and 5 zoonotic isolates, from China (26), Great Britain (27), Korea (10,12,13) and the U.S (28–30).

To assign genetic relationships within each taxon, we adopted the following hierarchical categories that represent the *least* to *most* genetic similarity as previously described (31–33). *Clade* denotes isolates that are monophyletic, *cluster* denotes isolates from different patients that are within a clade and within the SNP threshold of recent common ancestry (33). Isolates within a clade with SNP differences exceeding this threshold are considered *unclustered*. To identify clusters between patients, we applied a 20 SNP threshold (Figure 4, <https://wwwnc.cdc.gov/EID/article/27/11/21-0124-F4.htm>) and a 10 SNP threshold (Appendix 2 Figure 7).

Network visualization of relationships between patients with clustered isolates was performed with Cytoscape 3.8.1 (34). Clades of epidemiologically relevant clusters with three or more isolates, collected from two or more patients treated in the same facility were visualized using the MRCA function in the ggtree (35).

Treegubbins (4) was used to test for the presence of statistically-significant ( $p = 0.01$ ; 1000 permutations) increases in node density within the *M. avium* and *M. intracellulare* subspecies *chimaera* and *intracellulare* phylogenies. Clades with significant increases in node density were examined for the number of: i) isolates; ii) persons with cystic fibrosis (CF); iii) non-RDP isolates. Clades containing isolates from only one patient were excluded from further designations.

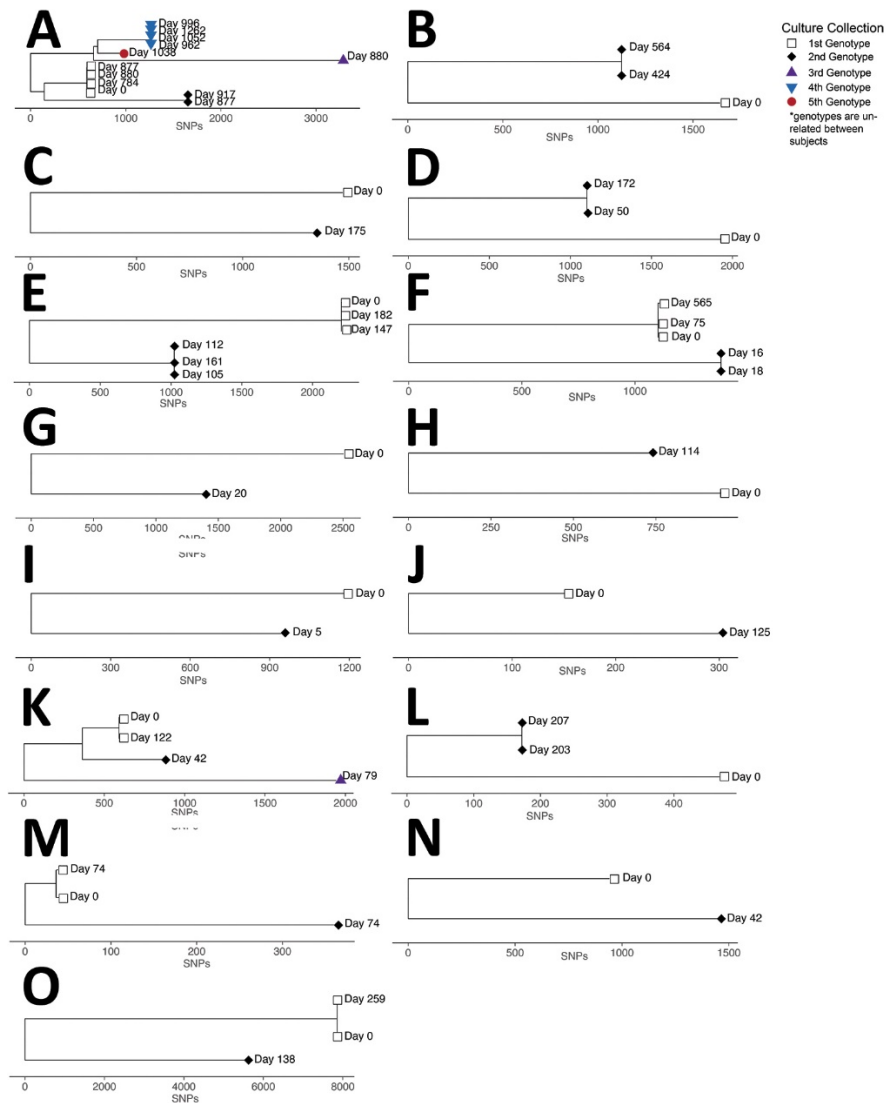
## References

1. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol*. 2017;13:e1005595.
2. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
3. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. *SNP-sites*: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom*. 2016;2:e000056.
4. Harris S. TreeGubbins. 2016 [cited yyyy mmm d]. [https://github.com/simonrharris/tree\\_gubbins](https://github.com/simonrharris/tree_gubbins)
5. Yano H, Iwamoto T, Nishiuchi Y, Nakajima C, Starkova DA, Mokrousov I, et al. Population structure and local adaptation of MAC lung disease agent *Mycobacterium avium* subsp. *hominissuis*. *Genome Biol Evol*. 2017;9:2403–17.
6. Uchiya K, Takahashi H, Yagi T, Moriyama M, Inagaki T, Ichikawa K, et al. Comparative genome analysis of *Mycobacterium avium* revealed genetic diversity in strains that cause pulmonary and disseminated disease. *PLoS One*. 2013;8:e71831.
7. Sanchini A, Semmler T, Mao L, Kumar N, Dematheis F, Tandon K, et al. A hypervariable genomic island identified in clinical and environmental *Mycobacterium avium* subsp. *hominissuis* isolates from Germany. *Int J Med Microbiol*. 2016;306:495–503.
8. Lahiri A, Kneisel J, Kloster I, Kamal E, Lewin A. Abundance of *Mycobacterium avium* ssp. *hominissuis* in soil and dust in Germany—implications for the infection route. *Lett Appl Microbiol*. 2014;59:65–70.
9. Bruffaerts N, Vluggen C, Duytschaever L, Mathys V, Saegerman C, Chapeira O, et al. Genome sequences of four strains of *Mycobacterium avium* subsp. *hominissuis*, isolated from swine and humans, differing in virulence in a murine intranasal infection model. *Genome Announc*. 2016;4:e00533–16.
10. Quan TP, Bawa Z, Foster D, Walker T, Del Ojo Elias C, Rathod P, et al.; MMM Informatics Group. Evaluation of whole-genome sequencing for mycobacterial species identification and drug susceptibility testing in a clinical setting: a large-scale prospective assessment of performance against line probe assays and phenotyping. *J Clin Microbiol*. 2018;56:e01480–17.

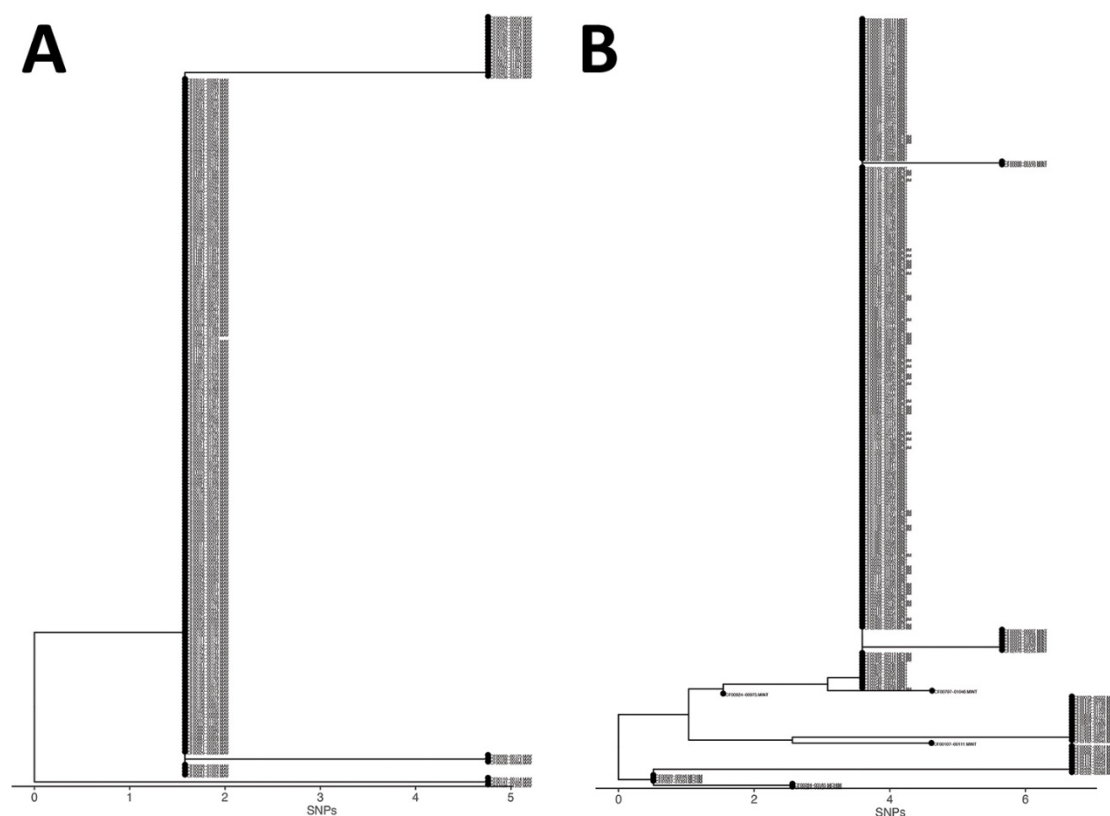
11. Bryant JM, Thibault VC, Smith DG, McLuckie J, Heron I, Sevilla IA, et al. Phylogenomic exploration of the relationships between strains of *Mycobacterium avium* subspecies *paratuberculosis*. BMC Genomics. 2016;17:79.
12. Chand M, Lamagni T, Kranzer K, Hedge J, Moore G, Parks S, et al. Insidious risk of severe *Mycobacterium chimaera* infection in cardiac surgery patients. Clin Infect Dis. 2017;64:335–42.
13. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, et al.; COMPASS-TB Study Group. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med. 2016;4:49–58.
14. Lande L, Alexander DC, Wallace RJ Jr, Kwait R, Iakhiaeva E, Williams M, et al. *Mycobacterium avium* in community and household water, suburban Philadelphia, Pennsylvania, USA, 2010–2012. Emerg Infect Dis. 2019;25:473–81.
15. Operario DJ, Pholwat S, Koepfel AF, Prorock A, Bao Y, Sol-Church K, et al. *Mycobacterium avium* complex diversity within lung disease, as revealed by whole-genome sequencing. Am J Respir Crit Care Med. 2019;200:393–6.
16. Matern WM, Bader JS, Karakousis PC. Genome analysis of *Mycobacterium avium* subspecies *hominissuis* strain 109. Sci Data. 2018;5:180277.
17. Zhao X, Epperson LE, Hasan NA, Honda JR, Chan ED, Strong M, et al. Complete genome sequence of *Mycobacterium avium* subsp. *hominissuis* strain H87 isolated from an indoor water sample. Genome Announc. 2017;5:e00189–17.
18. Horan KL, Freeman R, Weigel K, Semret M, Pfaller S, Covert TC, et al. Isolation of the genome sequence strain *Mycobacterium avium* 104 from multiple patients over a 17-year period. J Clin Microbiol. 2006;44:783–9.
19. Pfeiffer W, Braun J, Burchell J, Witte CL, Rideout BA. Whole-genome analysis of mycobacteria from birds at the San Diego Zoo. PLoS One. 2017;12:e0173464.
20. Li L, Bannantine JP, Zhang Q, Amonsin A, May BJ, Alt D, et al. The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. Proc Natl Acad Sci U S A. 2005;102:12344–9.
21. Mac Aogáin M, Roycroft E, Raftery P, Mok S, Fitzgibbon M, Rogers TR. Draft genome sequences of three *Mycobacterium chimaera* respiratory isolates. Genome Announc. 2015;3:e01409–15.

22. Hasan NA, Honda JR, Davidson RM, Epperson LE, Bankowski MJ, Chan ED, et al. Complete genome sequence of *Mycobacterium chimaera* strain AH16. *Genome Announc.* 2016;4:e01276–16.
23. Hasan NA, Lawsin A, Perry KA, Alyanak E, Toney NC, Malecha A, et al. Complete genome sequence of *Mycobacterium chimaera* strain CDC 2015–22–71. *Genome Announc.* 2017;5:e00693–17.
24. van Ingen J, Kohl TA, Kranzer K, Hasse B, Keller PM, Katarzyna Szafrńska A, et al. Global outbreak of severe *Mycobacterium chimaera* disease after cardiac surgery: a molecular epidemiological study. *Lancet Infect Dis.* 2017;17:1033–41.
25. Kim BJ, Kim BR, Lee SY, Seok SH, Kook YH, Kim BJ. Whole-genome sequence of a novel species, *Mycobacterium yongonense* DSM 45126T. *Genome Announc.* 2013;1:e00604–13.
26. Hasan NA, Warren RL, Epperson LE, Malecha A, Alexander DC, Turenne CY, et al. Complete genome sequence of *Mycobacterium chimaera* SJ42, a nonoutbreak strain from an immunocompromised patient with pulmonary disease. *Genome Announc.* 2017;5:e00963–17.
27. Zhang X, Lin J, Feng Y, Wang X, McNally A, Zong Z. Identification of *Mycobacterium chimaera* in heater-cooler units in China. *Sci Rep.* 2018;8:7843.
28. Kim B-J, Choi B-S, Lim J-S, Choi I-Y, Kook Y-H, Kim B-J. Complete genome sequence of *Mycobacterium intracellulare* clinical strain MOTT-64, belonging to the INT1 genotype. *J Bacteriol.* 2012;194:3268.
29. Kim B-J, Choi B-S, Lim J-S, Choi I-Y, Lee J-H, Chun J, et al. Complete genome sequence of *Mycobacterium intracellulare* clinical strain MOTT-02. *J Bacteriol.* 2012;194:2771.
30. Kim B-J, Choi B-S, Lim J-S, Choi I-Y, Lee J-H, Chun J, et al. Complete genome sequence of *Mycobacterium intracellulare* strain ATCC 13950(T). *J Bacteriol.* 2012;194:2750.
31. Hasan NA, Epperson LE, Lawsin A, Rodger RR, Perkins KM, Halpin AL, et al. Genomic analysis of cardiac surgery–associated *Mycobacterium chimaera* infections, United States. *Emerg Infect Dis.* 2019;25:559–63.
32. Perkins KM, Lawsin A, Hasan NA, Strong M, Halpin AL, Rodger RR, et al. Notes from the field: *Mycobacterium chimaera* contamination of heater-cooler devices used in cardiac surgery—United States. *MMWR Morb Mortal Wkly Rep.* 2016;65:1117–8.

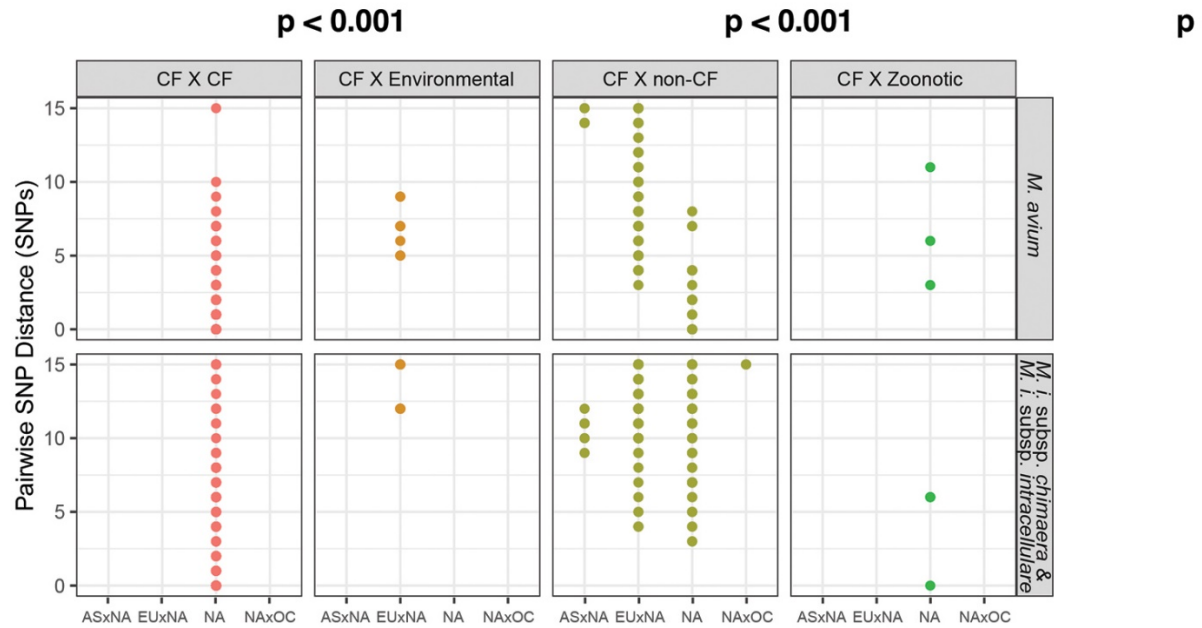
33. Davidson RM, Hasan NA, Epperson LE, Benoit JB, Kammlade SM, Levin AR, et al. Population genomics of *Mycobacterium abscessus* from United States cystic fibrosis care centers. *Ann Am Thorac Soc*. 2021 Apr 15 [Epub ahead of print].
34. Bryant JM, Grogono DM, Rodriguez-Rincon D, Everall I, Brown KP, Moreno P, et al. Emergence and spread of a human-transmissible multidrug-resistant nontuberculous mycobacterium. *Science*. 2016;354:751–7.
35. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–504.



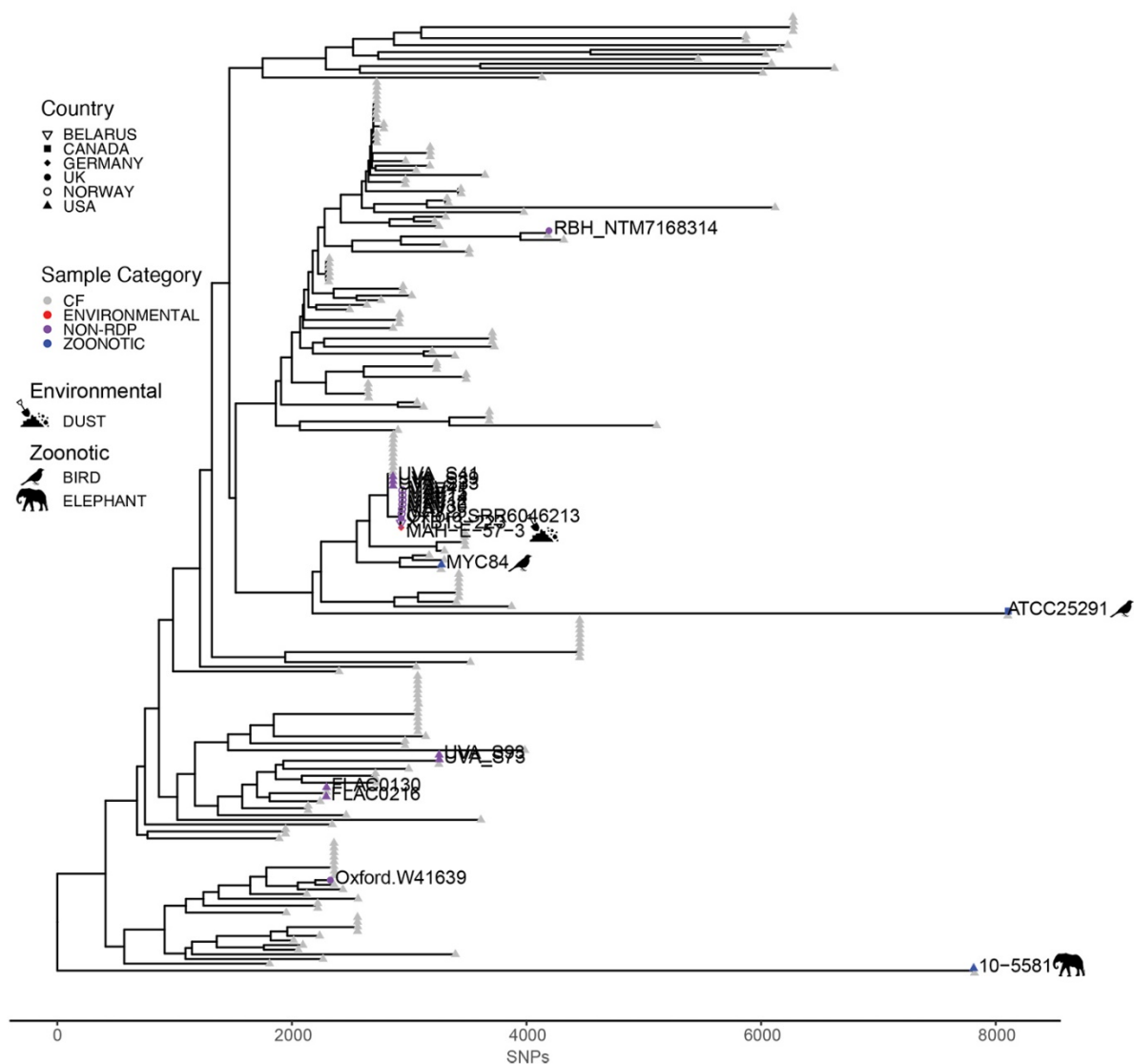
**Appendix 2 Figure 1.** Phylogenetic visualization of multiple *Mycobacterium avium* and *Mycobacterium intracellulare* subsp. *intracellulare* genotypes found in patients with two or more isolates sequenced. A. Five *M. avium* (MAV) genotypes in 12 isolates collected from patient CF00002. B. Two MAV genotypes in three isolates collected from patient CF00008. C. Two MAV genotypes in two isolates collected from patient CF00012. D. Two MAV genotypes in three isolates collected from patient CF00026. E. Two MAV genotypes in six isolates collected from patient CF00029. F. Two MAV genotypes in five isolates collected from patient CF00052. G. Two MAV genotypes in two isolates collected from patient CF00060. H. Two MAV genotypes in two isolates collected from patient CF00100. I. Two MAV genotypes in two isolates collected from patient CF00193. J. Two MAV genotypes in two isolates collected from patient CF00745. K. Three *M. MAV* genotypes in four isolates collected from patient CF00776. L. Two MAV genotypes in three isolates collected from patient CF00812. M. Two MAV genotypes in three isolates collected from patient CF00880. N. Two *M. intracellulare* subsp. *intracellulare* (MINT) genotypes in two isolates collected from patient CF00004. O. Two MINT genotypes in three isolates collected from patient CF00131.



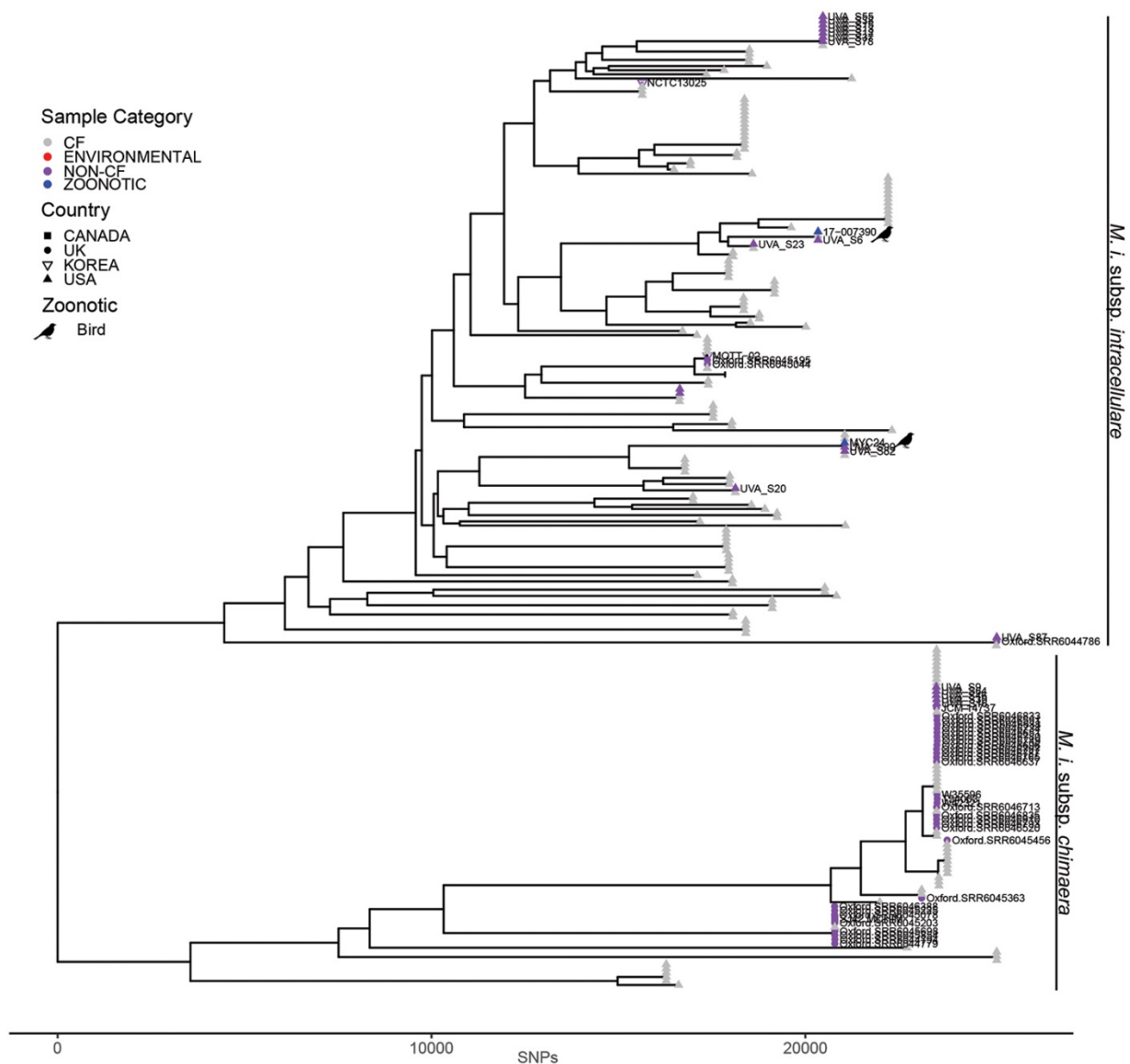
**Appendix 2 Figure 2.** Phylogenetic clustering of *rpoB* sequences from US CF MAC isolates. A. Phylogeny of 186 *M. avium* (MAV) isolates from 93 patients. B. Phylogeny of 44 *M. intracellulare* subsp. *chimaera* (MCHIM) and 134 *M. intracellulare* subsp. *intracellulare* (MINT) isolates' *rpoB* sequences. *rpoB* from each submitted isolate were analyzed using 250 bootstrap replicates of the observed SNPs and compared to identify phylogenetic clusters.



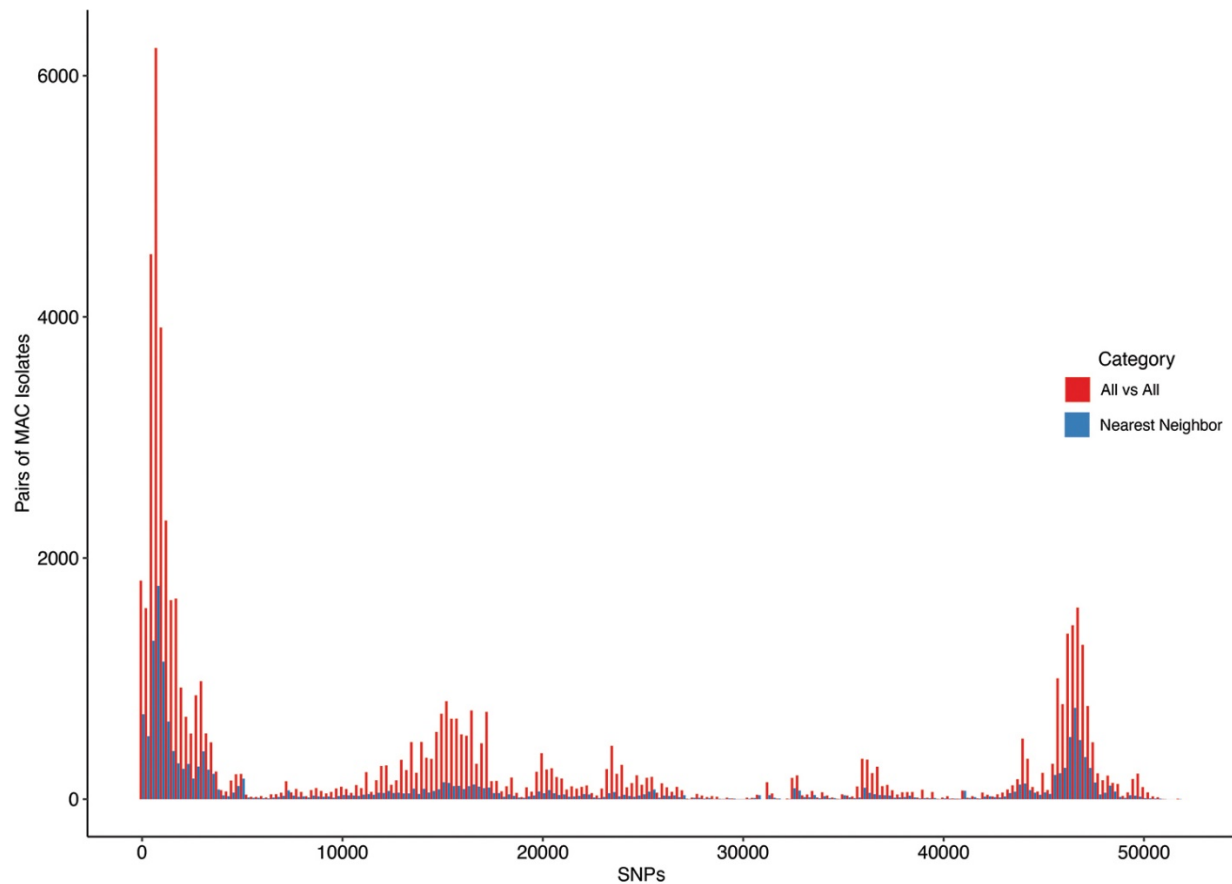
**Appendix 2 Figure 3.** Matches between US CF Care Center isolates from US persons with CF and environmental, non-CF clinical and zoonotic isolates from around the world. Each row represents the categorical comparisons for *M. avium* (MAV) or *M. intracellulare* subsp. *chimaera* (MCHIM) and *M. intracellulare* subsp. *intracellulare* (MINT). Comparisons between isolates of different continents were arranged along the x-axis are abbreviated for the continents of origin (i.e., Asia: AS; Europe: EA, North America: NA, Oceania: OC).



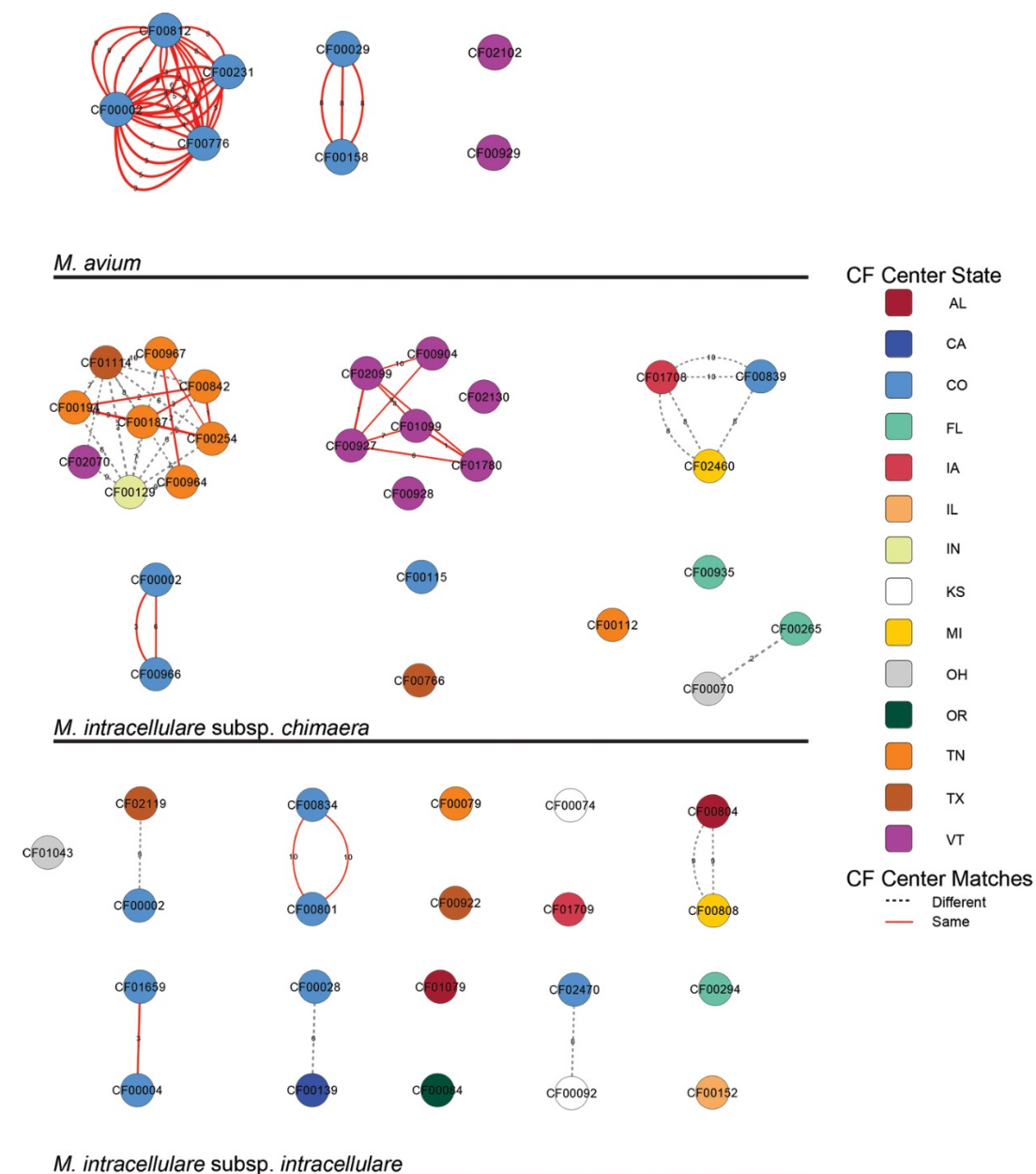
**Appendix 2 Figure 4.** Phylogenetic Visualization of *M. avium* isolates from US persons with CF and matches to non-CF, Environmental and Zoonotic isolates. Phylogenetic relationships between isolates from patients with CF (gray), environmental (red), non-CF (purple), and zoonotic (blue) isolates from Belarus (downward triangle), Canada (square), Germany (diamond), UK (solid circle), Norway (empty circle) and the USA (triangle) were visualized. Similar environmental isolate was dust-derived (dust symbol) and zoonotic isolates were bird-derived (bird symbol) and elephant-derived (elephant symbol). The scale bar along the x-axis represents the number of single nucleotide polymorphisms (SNPs).



**Appendix 2 Figure 5.** Phylogenetic Visualization of *Mycobacterium intracellulare subspecies chimaera* and *intracellulare* isolates from US persons with CF and matches to non-CF, Environmental and Zoonotic isolates. Phylogenetic relationships between isolates from persons with CF (gray), environmental (red), non-CF (purple), and zoonotic (blue) isolates from Canada (square), UK (circle), Korea (downward triangle) and the USA (triangle) were visualized. Similar zoonotic isolates were all bird-derived (bird symbol). The scale bar along the x-axis represents the number of single nucleotide polymorphisms (SNPs).



**Appendix 2 Figure 6.** Pairwise SNP distances of All vs All compared to Nearest Neighbors in MAC isolates infecting US persons with cystic fibrosis. The pairs of isolates (y-axis) having observed SNPs (x-axis) between all *M. avium* (MAV) and *M. intracellulare* subsp. *chimaera* (MCHIM), and *M. intracellulare* subsp. *intracellulare* (MINT) isolates from within the same patient and between different patients (All vs All; red) are compared to the nearest neighbor pair of isolates (blue). with the lowest SNPs observed between all possible pairings of within same patient and between patients is shown. The nearest-neighbor is the combination of two isolates with the lowest pairwise SNP difference observed between each possible pair of patients.



**Appendix 2 Figure 7.** Genetic clusters of *Mycobacterium avium*, *Mycobacterium intracellulare* subspecies *chimaera* and *intracellulare* in US persons with cystic fibrosis defined by a threshold of 10 SNPs or fewer between patients. Three clusters of *M. avium*, five clusters of *M. intracellulare* subsp. *chimaera* and 10 clusters of *M. intracellulare* subsp. *intracellulare* were identified. Each node represents a patient with at least one isolate having significant genetic similarity to an isolate in one or more patient(s). The color of each node represents the state of the submitting CF Care Center. Each edge represents genetic similarity between the isolates. Connecting edges are colored by matches within a CF Center (red) or between different CF Centers (dashed gray) and edge thickness is weighted from 0 SNPs (thickest) to 10 SNPs (thinnest) and the exact number of SNPs specified. Nodes with multiple connecting edges represent multiple isolates matching between the patients.